Reply to Reviewer 1:

1. There are a few typos in the paper. For example, it should be degrees of freedom instead of degree of freedom.

   We have corrected the typo you mentioned above. In addition, we have carefully went through the manuscript, and also asked an English native speaker to go through the manuscript.

2. Notations/definitions should be further clarified, especially when a symbol is first used, including α_m on pg 2 and β_0 on pg 3."

   Thanks for the comment. We pay more attention on the notation and definition, and make sure to pre-define the symbols for future usage. We have made revision on the symbols mentioned above including α_m on Page 2 and β_0 on Page 3.

3. On pg 7, there is a statement that α_i and ϵ_{ij} are independent I resume here you meant b_i, the random intercept, right?

   Yes, it should be b_i. We have corrected it on Page 7.

4. In the second data example, how was the variable time treated?

   The variable "time" is treated as a continuous variable. We updated our fitting model a little bit. In particular for the variables included in the model fitting, $Baseline$ is noted by the baseline epileptic seizure count rate per week; Trt is a binary indicator for treatment (1=progabide; 0=control); $Time$ is the number of weeks, which is valued by 2, 4, 6, and 8; $Interval\_duration$ is the duration of each interval, and $log(Interval\_duration)$ is treated as an offset variable in the model. We have updated the second paragraph on Page 8 and Table 4 as well as the corresponding results summary.

5. The simulation studies offered suggestions on applicable sample sizes for various variance estimators. However, how does the cluster size kick in? A more extensive discussion will be beneficial for practitioners.

   Thanks for the comment. In current simulation studies, we tried different scenarios with equal cluster size $n = 5, 10, 20$, and provided a general rule of appropriate sample sizes to preserve

Type I error for various variance estimators. To further investigate the effect of cluster sizes, we run additional simulations for the cases with binary outcomes and equal cluster size as well. For each variance estimator, the sample size $K = 10, 20, 30, 40, 50$ and a wider range of the cluster size $n = 5, 20, 50, 80, 100$ are investigated. We consider two correlation structures, independence and exchangeable, but the results are similar to each other. Thus, only the results using exchangeable correlation structure are provided and shown in Figure 1 and Table 1 below. From Figure 1, we can see that Type I error rates fluctuate around 0.05 varied by cluster size for each variance estimator with the recommended sample size in the manuscript. Also, from Table 1, we found out that the higher cluster size n can somewhat improve the performance in preserving Type I error, but the effect is not as substantial as the sample size K. In other words, when K is quite small, the performance on preserving Type I error is still not good even though n is extremely high. Please refer to the asymptotic properties of the parameter estimates in GEE (1). In addition, due to the fact that in most practical longitudinal designs, the cluster size (i.e., the number of observations within-subject) is usually less than 30 (2; 3). Thus, our recommendation can be applied in general cases (i.e., $n \geq 5$) based on current extensive simulations. We have made revision on the statements in the second paragraph on Page 7, and also add the limitation of our work in the first paragraph of Section 5 on Page 9.

Literature Cited

[1] Liang KY and Zeger SL. A Comparison of Two Bias-Corrected Covariance Estimators for Generalized Estimating Equations. *Biometrika* 1986;**73**: 13-22.

[2] Ma Y, Mazumdar M and Memtsoudis SG. Beyond repedated measures ANOVA: advanced statistical methods for the analysis of longitudinal data in anesthesia research. *Reg Anesth pain Med* 2012;**37(1)**: 99-105.

[3] Locascio JJ and Atri A. An overview of longitudinal data analysis methods for neurological research. *Dement Geriatr Cogn Discord Extra* 2011;**1**: 330-357.
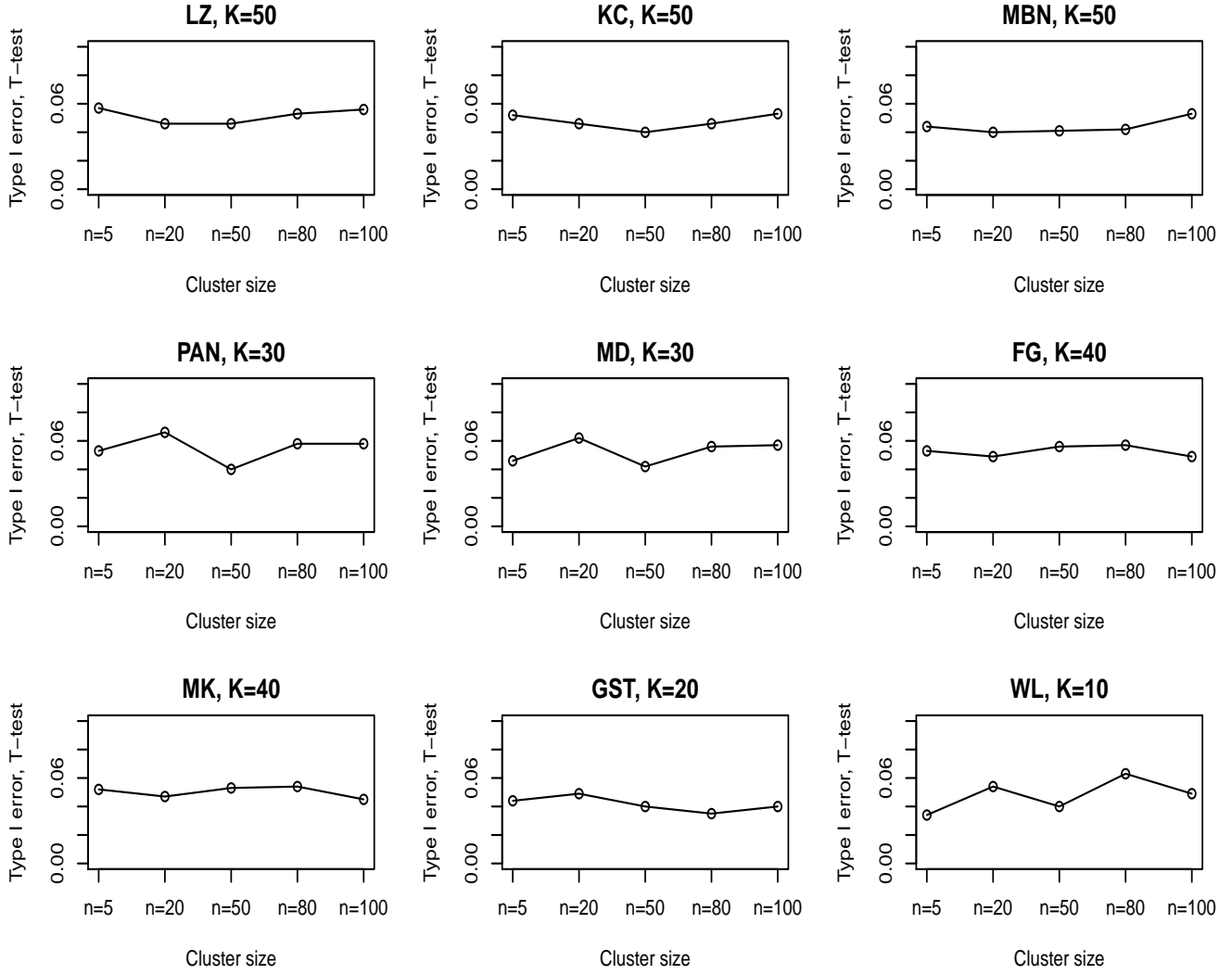
Figure 1: Type I errors based on $t-$tests for binary outcomes with the true correlation structure as exchangeable. Equal cluster sizes are considered for each scenario with the values of 5, 20, 50, 80, 100. The sample size K is the recommended value for perserving Type I error.

Table 1: Type I error for the case with binary outcomes based on $t-$tests

K		$n=5$	$n=20$	$n=50$	$n=80$	$n=100$
10	LZ	0.069	0.072	0.055	0.090	0.066
	MK	0.046	0.048	0.038	0.055	0.041
	PAN	0.047	0.055	0.044	0.064	0.056
	GST	0.031	0.030	0.026	0.041	0.024
	KC	0.040	0.047	0.038	0.053	0.039
	MD	0.039	0.047	0.038	0.056	0.043
	FG	0.058	0.055	0.048	0.071	0.054
	MBN	0.001	0.015	0.023	0.044	0.029
	WL	0.044	0.054	0.046	0.053	0.049
20	LZ	0.070	0.077	0.057	0.052	0.067
	MK	0.059	0.061	0.047	0.046	0.056
	PAN	0.055	0.058	0.054	0.052	0.058
	GST	0.044	0.049	0.040	0.035	0.040
	KC	0.056	0.056	0.053	0.044	0.055
	MD	0.046	0.062	0.042	0.056	0.057
	FG	0.065	0.066	0.054	0.050	0.061
	MBN	0.014	0.048	0.038	0.040	0.048
	WL	0.051	0.056	0.053	0.051	0.056
30	LZ	0.054	0.076	0.050	0.063	0.065
	MK	0.049	0.064	0.044	0.056	0.057
	PAN	0.053	0.046	0.050	0.058	0.048
	GST	0.046	0.056	0.033	0.045	0.045
	KC	0.048	0.068	0.041	0.051	0.054
	MD	0.052	0.060	0.048	0.046	0.056
	FG	0.049	0.071	0.046	0.060	0.058
	MBN	0.019	0.055	0.040	0.050	0.053
	WL	0.050	0.065	0.040	0.058	0.058
40	LZ	0.056	0.054	0.060	0.060	0.051
	MK	0.052	0.047	0.053	0.054	0.045
	PAN	0.052	0.047	0.049	0.055	0.048
	GST	0.044	0.039	0.039	0.050	0.041
	KC	0.054	0.047	0.054	0.053	0.047
	MD	0.049	0.046	0.053	0.054	0.045
	FG	0.053	0.049	0.056	0.047	0.049
	MBN	0.036	0.041	0.053	0.049	0.044
	WL	0.051	0.047	0.046	0.055	0.048
50	LZ	0.057	0.046	0.046	0.053	0.056
	MK	0.050	0.042	0.045	0.044	0.055
	PAN	0.050	0.045	0.045	0.043	0.053
	GST	0.045	0.041	0.040	0.036	0.049
	KC	0.052	0.046	0.050	0.046	0.053
	MD	0.050	0.042	0.044	0.044	0.055
	FG	0.054	0.044	0.045	0.049	0.055
	MBN	0.044	0.040	0.041	0.042	0.053
	WL	0.049	0.045	0.045	0.043	0.053

Note: 1) The exchangeable "working" correlation structure is considered; 2) The results of Type I error in red above are provided for each variance estimator under the scenario with the corresponding recommended appropriate sample size.